Hypertext as a Tool for Exploring Personal Data on Social Media

Eelco Herder
Radboud Universiteit Nijmegen
Institute for Computing and Information Sciences
Nijmegen, The Netherlands
eelcoherder@acm.org

Daniel Roßner
Hof University
Institute of Information Systems
Hof, Germany
daniel.rossner@iisys.de

Claus Atzenbeck
Hof University
Institute of Information Systems
Hof, Germany
claus.atzenbeck@iisys.de

ABSTRACT

Social networks such as Facebook are required to provide users with their personal data. However, these dumps do not provide users insight in overarching themes in their online behavior. In this poster, we discuss the development of *Mother*, a spatial hypertext system for visual data exploration. First insights include that the less obvious connections are more interesting and relevant to the user than very close semantic or temporal connections.

CCS CONCEPTS

• **Human-centered computing** → **Hypertext / hypermedia**; *Graphical user interfaces*; *User centered design*; • **Software and its engineering** → *Software infrastructure.*

KEYWORDS

spatial hypertext; social networks; user profiles; GDPR; data exploration

ACM Reference Format:
Eelco Herder, Daniel Roßner, and Claus Atzenbeck. 2020. Hypertext as a Tool for Exploring Personal Data on Social Media. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media (HT '20), July 13–15, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3372923.3404831

1 INTRODUCTION

Social networks are widely spread. The *Digital 2020* report[1] counts 2.5 billion active Facebook users, which makes Facebook the world's most-used social platform. The vast amount of information communicated by a huge group of people on a single social media platform creates significant advantages to the platform provider [2].

Increasingly, governments introduce regulations that protect citizens against the company's market power or lobbyists. The European General Data Protection Regulation (GDPR) [4] opens, among others, the right to EU citizens to request their data from platform providers.

In this poster, we present our system Mother [1] as a hypertext tool for exploring social media data. The goal is to increase the users' awareness of what data and, more importantly, which associations between information units the platform provider stores. We will use Facebook data dumps as an example for social media platforms.

[1]See https://datareportal.com/reports/digital-2020-global-digital-overview, data updated to 25 January 2020

HT '20, July 13–15, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7098-1/20/07.
https://doi.org/10.1145/3372923.3404831

2 BACKGROUND

Personal data is used by platforms for a variety of purposes, varying from personalization and recommendation to monetization, for example via advertisements and nudges to continue visiting the platform [5].

The introduction of the General Data Protection Regulation, the GDPR [4], in Europe provides end-users with means for requesting transparency. Dumps of personal data or posts from social media platforms may be voluminous. Therefore, they generally need to be teared apart into smaller coherent informational units and associated in a semantically meaningful way.

From its very beginning [3], hypertext has been been considered a medium for representing human associations, a medium for humans to express their (interconnected) ideas. A special type is *spatial hypertext* [6], which follows a *desk-on-table* metaphor: similar to physical paper notes on a desk, objects can be moved on a 2D canvas.

The big advantage of using spatial hypertext is that structures can be created and modified at ease and at a low cognitive load [7]. As such, spatial hypertext helps in exploring unknown knowledge by supporting creating, modifying, or destroying contexts of information to which the system reacts to.

3 SYSTEM DESIGN AND DEVELOPMENT

Following the GDPR regulations, Facebook allows its users to obtain a copy of their personal data as a simple download[2]. The user receives an archive with files that contain, among others, the user's own posts, comments, likes and reactions to posts by others (both friends and Facebook pages), search history, lists of friends, subscribed groups and pages, and interaction with advertisements.

The textual contents of the posts as well as post frequency statistics can provide rich material for users to obtain insight in and to reflect on their Facebook usage, including the reporting of life events, work-related announcements, discussions with friends, shared silly pictures or memes, and interaction with advertisements.

As a first step towards a visualization of Facebook data, we created a script to process the JSON Facebook post data of a user into a graph-based format, with posts, keywords, months and years as vertices, connected by edges with various weights. The keywords are extracted from the Facebook posts, converted into lowercase, lemmatized, stopwords removed and only keywords that appear in

[2]https://www.facebook.com/settings?tab=your_facebook_information

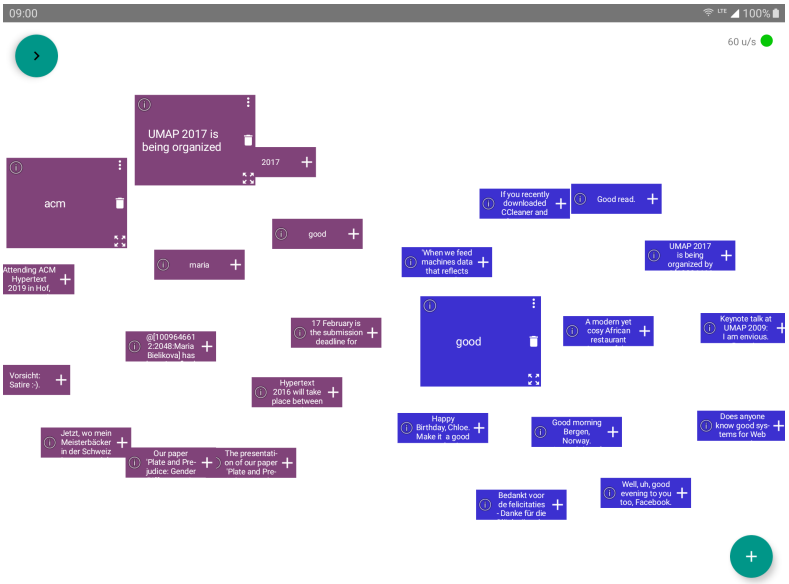

Figure 1: Screenshot of the prototype – one post and two keywords are added

at least 5 posts are stored, in order to keep the number within limits. Edges between posts, years, months and keyword were created and weighted based on tf-idf and/or co-occurrence.

From a user's point of view, the principle of Mother is that one selects a single entity (in our case, a Facebook post), which is then displayed along with *related entities* (other posts and keywords) as recommendations that can be followed in order to create a narrative. An example is shown in Fig. 1.

The first developed application area of Mother concerned the movie domain, where movies are connected with one another through actors, genres and other entities [1]. This content-based approach made sense in the movie domain, but applied to Facebook posts they resulted in relations that were too obvious for the user (e.g., birthday wishes were related to other birthday wishes). After all, there is a difference between exploring an unknown domain, and *introspection*, the examination or observation of one's own mental and emotional processes.

For this reason, we decided to only recommend posts based on a combination of content-based and temporal similarity (hour of day, day of week, month), manually tuned and evaluated by the authors in several sessions. Furthermore, we also added the keywords as vertices and related them to the posts. Given the large differences in posting behavior, even between the authors, it was concluded that no optimal a-priori values could be found. Instead, we opted for a configuration that led to the first author's observation that he typically posted his reflections on a day in the early evening; in addition, some randomness was added in order to prevent users to get locked in a small number of favorite themes.

4 CONCLUSIONS AND FUTURE WORK

An important lesson learned during the design of the system and exploration of the author's own profiles is that there is a difference between exploring an unknown domain (e.g., movies) and introspective exploration of one's own activities and posts: for unknown domains, close semantic relations (such as actor X plays in movie Y) are meaningful and useful, but when exploring one's own activities, these relations turn out to be to obvious to be useful.

As future work, we aim to improve the interface, so that the posts are presented along with photos and links, if present. Furthermore, we will use similar approaches for exploring different parts of a user's Facebook profile: we expect that users' remembrance of their own comments to friends' posts and page contents is far lower than of their own posts and therefore can provide more insight in which triggers one is sensitive to.

REFERENCES

[1] Claus Atzenbeck, Daniel Roßner, and Manolis Tzagarakis. 2018. Mother – An Integrated Approach to Hypertext Domains. In *Proceedings of the 29th ACM Conference on Hypertext and Social Media*. ACM Press, 145–149. https://doi.org/10.1145/3209542.3209570

[2] Joseph Bonneau, Jonathan Anderson, and George Danezis. 2009. Prying Data out of a Social Network. In *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*. 249–254. https://doi.org/10.1109/ASONAM.2009.45

[3] Vannevar Bush. 1945. As we may think. *The Atlantic Monthly* 176, 1 (7 1945), 101–108. http://www.theatlantic.com/doc/194507/bush

[4] European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* 59, L 119 (5 2016), 1–149. http://data.europa.eu/eli/reg/2016/679/oj

[5] Eelco Herder. 2019. The Need for Identifying Ways to Monetize Personalization and Recommendation. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. 291–294.

[6] Catherine C. Marshall, Frank G. Halasz, Russell A. Rogers, and William C. Janssen. 1991. Aquanet: a hypertext tool to hold your knowledge in place. In *Proceedings of the 3rd ACM Conference on Hypertext*. ACM, 261–275. http://doi.acm.org/10.1145/122974.123000

[7] Frank M. Shipman, J. Michael Moore, Preetam Maloor, Haowei Hsieh, and Raghu Akkapeddi. 2002. Semantics happen: Knowledge building in spatial hypertext. In *Proceedings of the 13th Conference on Hypertext and Hypermedia*. ACM Press, 25–34. http://doi.acm.org/10.1145/513338.513350